

Reference

NBS
Publi-
cations

NAT'L INST. OF STAND & TECH R.I.C.



A11105 131107

NBSIR 80-2053

Concepts of Model Confidence

Saul I. Gass
Lambert S. Joel

Operations Research Division
Center for Applied Mathematics
National Bureau of Standards
U.S. Department of Commerce
Washington, DC 20234

June 1980

Technical Report to:

Dr. George M. Lady

Office of Oversight Analysis and Access
Department of Energy
Washington, DC 20461

QC
100
U56
80-2053
1980

MAR 2 1981

100-2053-100

QC100

USC

NO. 80-2053

1980

NBSIR 80-2053

CONCEPTS OF MODEL CONFIDENCE

Saul I. Gass
Lambert S. Joel

Operations Research Division
Center for Applied Mathematics
National Bureau of Standards
U.S. Department of Commerce
Washington, DC 20234

June 1980

Technical Report to:
Dr. George M. Lady
Office of Oversight Analysis and Access
Department of Energy
Washington, DC 20461



U.S. DEPARTMENT OF COMMERCE, Philip M. Klutznick, Secretary

Luther H. Hodges, Jr., Deputy Secretary

Jordan J. Baruch, Assistant Secretary for Productivity, Technology, and Innovation

NATIONAL BUREAU OF STANDARDS, Ernest Ambler, Director

7-17-72

100-100000-100000

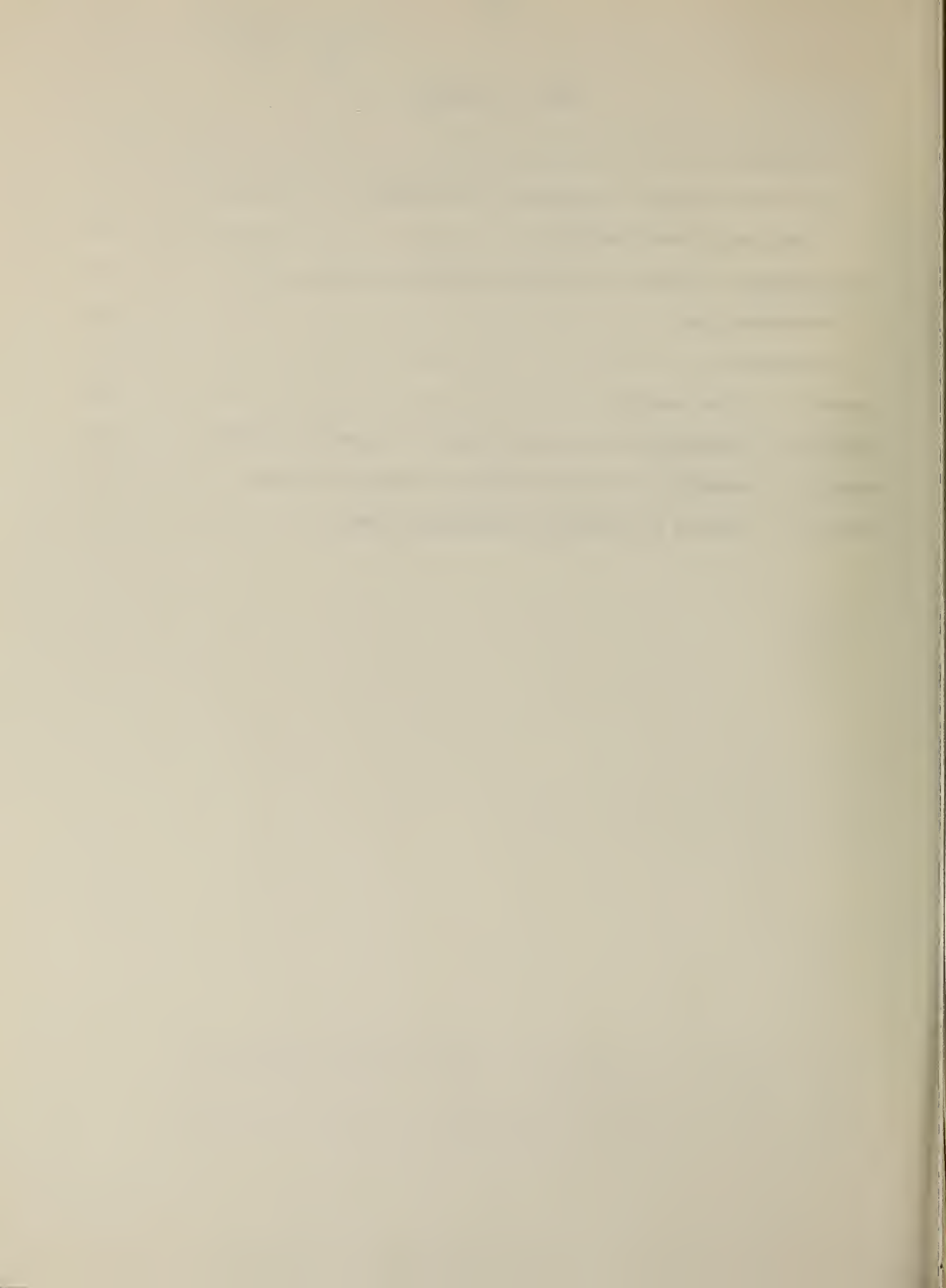
100-100000-100000
100-100000-100000
100-100000-100000

100-100000-100000

100-100000-100000
100-100000-100000
100-100000-100000
100-100000-100000

TABLE OF CONTENTS

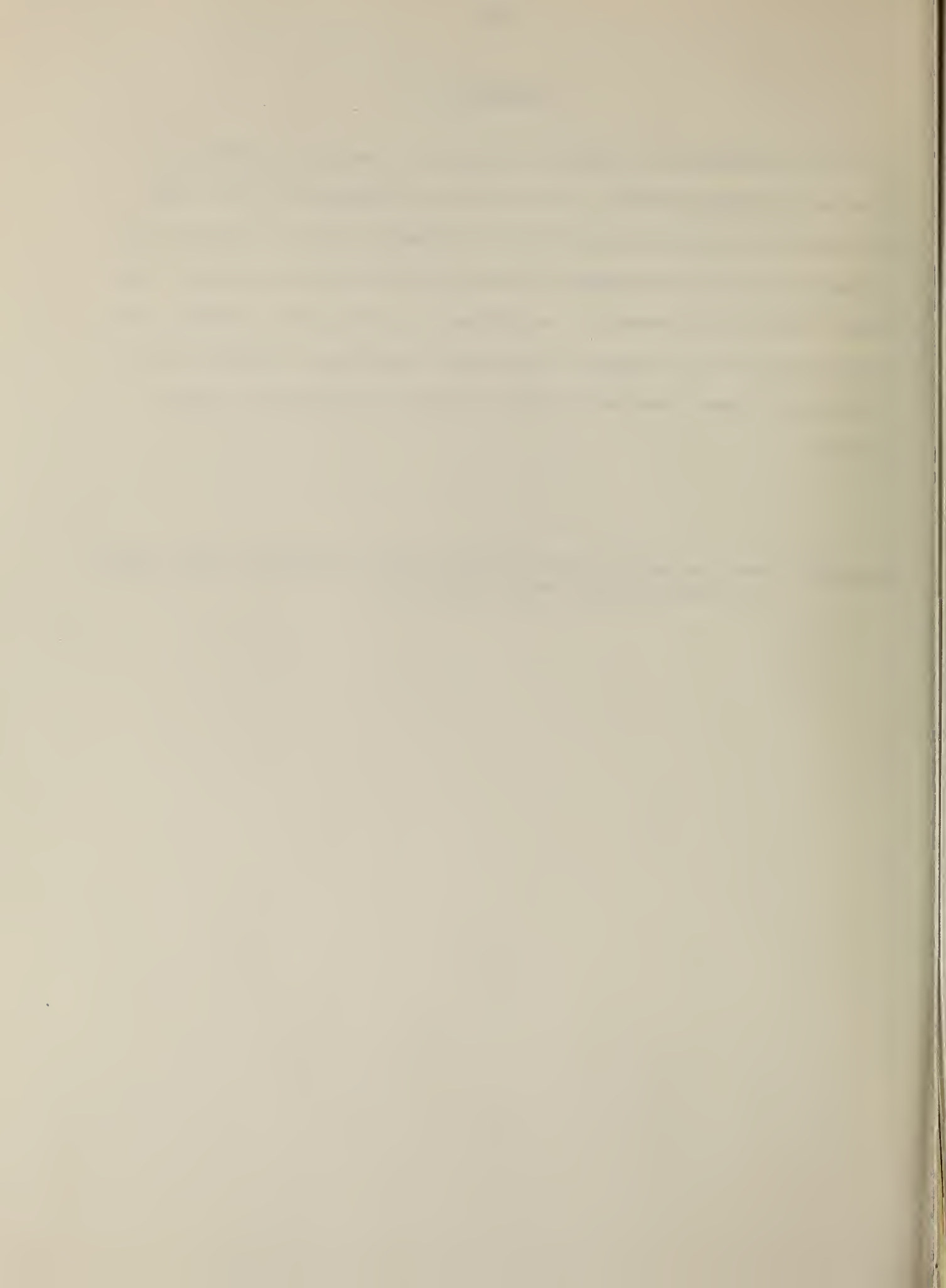
I. Introduction.....	1
II. The Decision Maker, the Analyst, and the Model.....	4
III. Establishing Model Confidence.....	8
IV. Confidence Criteria and the Model Evaluation Process.....	11
V. Recommendations.....	18
VI. References.....	19
Appendix A. Survey Results.....	22
Appendix B. Respondents from National Bureau of Standards Workshop.....	27
Appendix C. Respondents from NATO Brookhaven Energy Conference.....	29
Appendix D. Combined NBS/Brookhaven Respondent Totals.....	31



ABSTRACT

This report discusses the concept of confidence in results obtained from large-scale modeling systems. It is written in satisfaction of the "model confidence" tasks of a National Bureau of Standards project on "Energy Model Validation Procedure Development," funded by the Department of Energy. This report includes discussions of: our efforts to define model confidence; the workshop held for this purpose; a preliminary methodology to measure confidence; and a survey conducted to obtain opinions on significant related issues.

Key Words: Decision making; model assessment; model confidence; model evaluation; model utility; model validation.



I. INTRODUCTION

This report from the National Bureau of Standards (NBS) project* for "Energy Model Validation Procedure Development" is written in response to the following tasks from the scope of work.

Task 5: A specification of alternative concepts of "confidence" in system results will be prepared.

Task 6: A determination will be made of the relationship between the outcome of the various system attribute evaluations and the concepts of confidence. To the extent possible a rigorous statement of this relationship will be achieved.

Task 7: A summary concept of system result confidence will be developed to include the specification of the evaluation activities necessary to support the determination of system result confidence.

Task 8: An end of year report will be prepared on standards and procedures for determining system confidence.

These tasks are part of the NBS project for the Department of Energy (DOE) that has as its major goal the development of system validation procedures and their application to the latest version of the Midterm Oil and Gas Supply Modeling System.

*Sponsored by the Department of Energy, Office of Analysis Oversight and Access, Interagency Agreement No. EA77-A-01-6610.

The project's model confidence activities have taken the following form:

- (1) Development of criteria and measures of confidence;
- (2) Preparation of a discussion paper on model confidence;
- (3) The convening of a workshop to (a) define model confidence, (b) review current research relevant to the concept of model confidence, (c) discuss a preliminary methodology to be used to measure confidence, and (d) indicate areas of future research; and
- (4) An informal survey to obtain other opinions on significant issues related to model confidence. (The results and interpretation are given as an appendix to this report.)

Based on our review, it is apparent that a universal definition of model confidence does not exist. Past research does not include an operational approach that can be used by DOE to establish a concept of confidence. Thus, in what follows, we are led to present our assumptions relative to a decision maker's confidence in a model, give a brief overview of relevant past research, and offer a set of model confidence criteria and a process for measuring whether or not the criteria are met.

We note that our conclusions in this paper are tentative. Our recommendations on future model confidence research are limited to those basic activities that we feel will be of most benefit to DOE (see Section V). However, we wish to

point out that the problem of establishing confidence in a policy model is of concern to the modeling community at large [41]. Efforts to resolve this problem are well justified.

II. THE DECISION MAKER, THE ANALYST, AND THE MODEL

The use of a mathematical model as an aid for resolving a specific decision problem requires, on the part of the decision maker, some basis for accepting the model outputs as an active part of the decision information set. The role of model outputs in the decision process is based on the decision maker's understanding and evaluation of the total modeling process that has produced the outputs. Usually, the model outputs are modified and factored into an explicit or intuitive conceptual model of the decision maker. In an extreme case, the model can be allowed to define the decision. For decision makers, their confidence in a model is expressed by the influence the model's outputs had in the decision.

The phrase "model confidence" has a familiar and comforting ring to those involved in the development or use of decision (aiding) models. In general, one has an intuitive notion of what model confidence implies. When asked for a formal definition, we find its meaning is discovered to be felt rather than known. Some may think that "confidence" is a quality of a model and a rough equivalent of validity. We emphasize model confidence not as an attribute of a model, but of the model user. Thus, in this report, confidence will be considered from the point of view of the decision maker/user of models, rather than that of the analyst/developer, under the assumption that they differ. Model confidence is an expression of the user's total attitude toward the model and of the willingness to employ its results in making decisions.

Our approach requires us to differentiate between confidence in outputs and utility of the model. Utility denotes the usefulness of the model to the decision maker and involves confidence. Utility is concerned with the total operating milieu of the model.

Although a number of studies have identified what other authors feel are the determinants (criteria) of model utility [3, 5, 25, 32, 33, 34], there is very little recorded information as to why a specific decision maker used or did not use the outputs of a model. If the decision maker is a part of the model development team, then all other things being equal, the outputs are usually treated with a reasonably high level of confidence; but a rationale for use beyond pride of authorship needs to be established. On the other hand, a model used successfully by one individual may be given little weight by a new decision maker unless materials are presented that provide a sense of confidence to the new user.

The basic decision situation involves a single decision maker who has, at a minimum, an internal or mental model of the process being investigated. As the problem is studied, with a model or by other means, additional information is furnished to the decision maker. Somehow the decision maker weighs the information that is gathered and makes a decision. Let us hypothesize the situation in which the decision maker has a mental model (whatever it may be) and also a formal decision model whose outputs can be used as an aid in arriving at a decision. Without the latter model, the decision maker would make a decision according to the mental model. How and why does the decision maker modify the mental model solution as new information is produced using the decision model? On what basis are the decision model's results ignored? These questions are another way of asking "How are the various sources of information weighed and emphasized by the decision maker?"

We cannot address these questions directly at this time. We will, however, formulate a "rational procedure" that the decision maker can use to estimate the utility of a model.

For the decision maker's environment, we can consider two situations. The first involves a mental model and a newly developed decision model; the second involves a mental model and an established decision model (with a history of use). In each case, we are concerned with the materials describing the model and how these materials are interpreted by the decision maker in establishing model confidence. For the new decision model, an initial confidence level would usually be hard to fix. The decision maker may act based on a determination of how well the model satisfies implicit or stated criteria for that model in the given decision environment, e.g., are the results consistent with intuition. As a model is used, a record and analysis of its results will enable the decision maker to adjust the estimate of confidence. (Long-term use of a model should not be prima facie evidence of a high degree of confidence by other users; many "imbedded" models are used habitually by an organization without any current justification.)

Confidence in a model is a result of the accumulation of information, the sum total of which leads to a judgmental statement by the decision maker. The generation of this information--what we term the model materials or documentation--is the task of the model analyst and developers. Some of this material will be produced to satisfy the needs and requests of the decision maker. The materials furnished should enable the decision maker to evaluate the model vis-a-vis any formal or informal criteria used to establish a measure of confidence. Not to produce the materials represents a failure in the model development process.

For the models of interest--decision models that are used as aids in determining policy--the level of confidence can vary from user to user because of differences in application requirements, as well as subjective judgmental preferences. Confidence in a model evolves by a joint effort between the model developers and a designated user. We can take an extreme position by saying that a decision model without a designated user (which implies a specific use) has no basis upon which a confidence statement can be made, i.e. the a priori confidence level is zero. Certainly many analysts can demonstrate that their models give quite accurate predictions, and thus the analysts have a high degree of confidence in the outputs. But for such models to be used in specific decision settings, the results must be evaluated in terms of the decision makers' criteria.

III. ESTABLISHING MODEL CONFIDENCE

It is clear that there is no single measure of model confidence or no absolute claim concerning the confidence that can be given a model or its output. For all but the simplest of decision models, we cannot expect to obtain statistical or numerical bases for statements of confidence. The situation is analogous to determining the confidence given to an expert witness in court. The judge and jury use criteria, usually of a qualitative nature, to determine the extent to which they let the expert's testimony influence their decision. A decision maker is a judge faced with an expert witness--the analyst or model developer--who has a magic black-box of a model in the computer room. Sometimes, the reputation and presentation of the witness are assumed to be sufficient reason to accept the testimony. But the astute decision maker (or the astute Congressman) no longer is satisfied with the outputs unless model confidence has been established in terms of the decision maker's criteria. What are these criteria? What form should they take? How consistent are they between decision makers and models? Given explicit criteria, how can an investigator "measure" a model's material to determine if the criteria are met? These are difficult questions to answer. In what follows, we shall offer an initial approach to resolving these questions and outline areas for further research.

A number of researchers have investigated the problems of evaluation, assessment, validation, credibility, reliability, and related model concerns. Many have described approaches that relate to the basic issue of establishing confidence in a model. These approaches, in general, employ loose definitions of

what is being investigated, be it an evaluation, statements of credibility, validity, etc.¹

We shall not review this material here (the reader is referred to [3, 4, 5, 8, 9, 25, 26, 32, 34, 36]) except to repeat the Professional Audit Review Team's (PART) recommendation [25] to DOE/EIA concerning procedures and practices for model building (note their use of credibility):

"To fulfill the intent of the Congress, we believe that EIA must establish the credibility of its mathematical and statistical models. In the 1977 PART report, we suggested the following procedures and practices as essential to building an acceptable level of credibility into EIA modeling activities.

1. Public Participation and Professional Review -- Outside professionals should be involved in the development and maintenance of a model, thus guaranteeing its widespread acceptance and credibility. Such involvement should include procedures that allow (1) internal and outside experts to participate in determining, updating, and refining major changes in assumptions and structure and (2) the general public to review and comment on the model's assumptions and structure.

2. Control over Model Changes -- A systematic procedure should exist that specifies what, when, and why changes should be made to the model and who should make them. This should take the form of a timetable for selected changes, a public list of individuals responsible for making changes, and a schedule of regular and planned uses of the model.

3. Documentation -- During the design, development, and maintenance of a computer model, its purpose, methodology, assumptions, capabilities, and limitations must be recorded and explained. An adequately documented model permits outside parties to use and understand it, evaluate its credibility, and participate in its development.

4. Verification -- To achieve credibility, a model's mathematical calculations should be checked for accuracy. Also, its structure and relationships should be verified against the system it is trying to represent.

¹We prefer the word confidence in that a claim of confidence in a model implies the intention to use the model. Confidence also implies credibility (believable, plausible, and worth of trust) and reliability (dependable), where credibility and reliability are attributes that can be measured only after the model has been used. Sargent [38] discusses the credibility of a modeler or institution and "confidence" in its models.

5. Validation -- A model's predictions should be compared with actual data to determine the probability of error in forecasts. This should be done on a regular basis with the results made available to the public.

6. Sensitivity Testing -- The extent that a model responds to changes in assumptions, specifications, and data should be measured. Again, the results of such tests should be made public."

If these procedures are followed during model building/development, then the task of model assessment is greatly reduced, requiring essentially no more than a review of the modeling process and selective testing. On the other hand, if these tasks have not been (well) executed by the modelers, their accomplishment falls on model assessors. In either case, completion of the procedures described above should be a major step in establishing model credibility and instilling confidence in model users.

IV. CONFIDENCE CRITERIA AND THE MODEL EVALUATION PROCESS

Our research approach to model confidence follows the basic directions given in the Scope of Work, i.e., determine how the evaluation of system (model) attributes relates to the concept of model confidence. Task 3 of the Scope of Work cited the following system attributes (as a minimum) to be evaluated:

- Completeness and accuracy of underlying data
- Conceptual sufficiency of system specification
- Appropriateness of operating representation
- Appropriateness of embodied estimation methodologies
- System sensitivity and stability
- System performance compared to known outcomes
- Computer related system characteristics
- Any other system element or attribute which significantly influences the confidence in system results.

In this section, we list our set of criteria that relate to model confidence. These criteria are not necessarily of a quantitative nature. Whether a model satisfies a criterion depends on the analysts' (or assessors') ability to produce specific information required by the decision maker. The ideal situation has the decision maker and analysts agreeing to criteria and information needs prior to and during model development and testing. The final model materials should then include the necessary information or explain why such information is unobtainable. A similar process should be part of any model evaluation, since such evaluations make sense only if they are done for designated decision problems and hence, for an implied set of decision makers.

Our discussion of decision-maker confidence and model evaluation cannot and should not be equated to any form of model certification. Our conclusion is that model confidence is a personal affair, with each decision maker internalizing the available information by means of an imprecise algorithm for evaluating model confidence. However, we do feel that such algorithms should be based on commonly accepted professional practices that can be expressed to a useful degree by information produced by the initial model analysts or by subsequent model assessors. It is our purpose here to detail such information requirements and then illustrate an approach that can be used by a decision maker to obtain statements of model confidence.

There are a number of ways to group the information requirements. We shall use one that is rather aggregated, recognizing that each heading can be expanded into subheadings. Our rationale for a restricted set of headings is that any measure of model confidence is based on many attributes and the mental process of converting corresponding information into a single measure is simplified if there are fewer elements to be considered. A more detailed approach is given in [3,4,40]. An item for future research is to determine which information is of importance to a decision maker. Our assumption is that a decision maker will review the information to determine the extent to which it satisfactorily addresses the topic with respect to a particular problem setting. The topics are the criteria on which model confidence will be judged.

A. Confidence Criteria for a Model

1. Model Definition--the problem and model environments: includes identification of the decision problems and related questions that the model is intended to address; and describes any prior use of the model to specific policy questions. The information gathered here should enable the decision maker to determine if the problem area in question is at least within the scope of the model purposes.
2. Model Structure--the theoretical and methodological bases of the model: includes assumptions required to fit the theory to the problem; and examination of methodologies and their assumptions, and the resultant model's appropriateness and applicability to specific problems. This information should enable the decision maker to determine if the model structure has limitations that preclude its use as a decision aid for the problem area in question.
3. Model Data--the data base, data sources, and procedures for data transformations: includes assumptions on representativeness and impartiality of data, how values of missing data are imputed, and data collection and audit procedures. This information should enable the decision maker to determine if data for the problem area in question are available at reasonable cost, are accurate enough, and are used correctly by the model.

4. Computer Model (Program) Verification--the tests and procedures used to debug the subprograms and program, and how the consistency between the program and model's mathematical and logical description was established. This information should enable the decision maker to determine if the computer program is reliable and if it appears to be an acceptable representation for the model.
5. Model Validation--methods by which the computer model has been analyzed in terms of its ability to produce results that can be relied upon by the decision maker: includes discussions on whether outputs are consistent with expected outcomes; comparisons with available historical results; analyses of sensitivity of key parameters; robustness and range of applicability of the model. This information should enable the decision maker to determine that the model's real-world approximation is suitable for the problem area in question.
6. Model Usability--resources, procedures, documentation, accessibility, transferability, and maintenance aspects of the model. This information should enable the decision maker to determine if the model can be used within the decision maker's problem environment.

7. Model Demographics--an abstract and description of the model antecedents and developmental process, originators and developers, past users, cost, and current developmental activities. This information should enable the decision maker to determine the model's status with respect to past achievements, theoretical and methodological state-of-the-art, and the expert advice that went into its development.

B. An Approach to Determining Model Confidence

Although at this time we cannot offer a universally acceptable measure of model confidence, we think that the information presented above can be utilized in the following approach to obtain statements of model confidence. Suppose the decision maker is furnished model evaluation material organized under the headings presented above. The decision maker implicitly forms some basis for reviewing the materials and determining what is required to state that a criterion is satisfied at a specified level. We shall assume a five-level structure for a criterion, with each level being characterized by a descriptive statement of opinion. We illustrate the approach and five levels using the "model validation" criterion. Five statements are constructed concerning model validation that indicate a sense of low to high confidence in this attribute. For example, on a scale of one (low) to five (high), the statements associated with model validation might be the following.

1. The validity of the model has not been demonstrated satisfactorily for the original problem environment.

2. The validity of the model has been demonstrated satisfactorily for the original problem environment, but there is some question as to whether the model will exhibit the same sense of validity for the new problem.
3. The model satisfies a minimal level of validity for the new problem; improvements are judged to be limited only by state-of-the-art.
4. Specific tests have indicated that the model will yield valid results for the new problem under a representative set of scenarios.
5. Specific tests, expert opinion and/or historical data indicate that the model will yield valid results for the new problem under a full range of reasonable scenarios. The model satisfies the criterion to the fullest extent possible.

Similar statements for the other criteria that represent the opinion of the decision maker would indicate the level at which each criterion was satisfied. We feel that five statements should be enough to capture the range of "not satisfying" to "fully satisfying" a criterion.

The presentation of the results can be done by using a bar chart approach that captures the interrelationships of all the criteria. We suggest something like Chart 1. The heavy lines in Chart 1 indicate the threshold boundaries of the criteria. That is, for a model and a given decision environment,

CRITERION	SCALE				
	1	2	3	4	5
Definition					
Structure					
Data					
Verification					
Validation					
Usability					
Demographics					

Chart 1

the decision maker, possibly in conjunction with the analysts or assessors, agrees to set a threshold value for each criterion. If the scale value falls below the threshold, then the model confidence in that area is in question. In the example, the levels judged achieved by the model are indicated in gray. Thus, this model meets the decision maker's minimum standard for "structure" and "usability," exceeds those for "definition," "verification," and "demographics," and fails to meet the standards for "data" and "validation."

V. RECOMMENDATIONS

As a research topic, the area of model confidence is an extensive one. It requires not only quantitative modeling talent, but expertise from other disciplines such as the behavioral and social sciences. What would benefit DOE the most, assuming a desire to continue a limited activity, is to expand upon the beginnings given in the preceding section. This can be done by performing the following research efforts:

- A. DOE continue the research in confidence by sponsoring a task that develops criteria and related statements from the perspective of DOE and other government decision makers.
- B. A parallel effort should experiment with the organization of materials from a DOE model assessment project into sets of information that can be used by a decision maker to measure the seven criteria and test the confidence methodology proposed in this report.
- C. Design a confidence experiment in which a new DOE model is developed to include the decision makers and a confidence determination procedure.

VI. REFERENCES

1. Fishman, G. S. and Kiviat, P. J., "Digital Computer Simulation: Statistical Considerations," Rand Report. RM-5387-PR, The Rand Corp., Santa Monica, CA, 1967.
2. Greenberger, M. and Richels, R., "Assessing Energy Policy Models: Current State and Future Directions," Annual Review of Energy, Vol. 4, 1979.
3. Gass, S. I., "Evaluation of Complex Models," Computers and Operations Research, pp. 27-35, Vol. 4, 1977.
4. Gass, S. I., "A Procedure for the Evaluation of Complex Models," Proceedings First International Conference in Mathematical Models, U. of Missouri, 1977.
5. U. S. General Accounting Office, "Guidelines for Model Evaluation," PAD-79-17, U. S. GAO, Washington, DC, January 1979.
6. House, P. W. and McLeod, J., Large-Scale Models for Policy Evaluation, John Wiley and Sons, NY, 1977.
7. Clark, J. and Cole, S., Global Simulation Models: A Comparative Study, John Wiley and Sons, NY, 1975.
8. U. S. General Accounting Office, "Review of the 1974 Project Independence Evaluation System," OPA-767-20, U. S. GAO, Washington, DC, 4/21/76.
9. U. S. General Accounting Office, "An Evaluation of the Use of the Transfer Income Model--TRIM--to Analyze Welfare Programs," PAD-78-14, U. S. GAO, Washington, DC, 1975.
10. Fromm, G., Hamilton, W. L., and Hamilton, D. E., "Federally Supported Mathematical Models: Survey and Analysis," U. S. GPO, Stock No. 038-000-0021-0, Washington, DC, 1975.
11. Shubik, M. and Brewer, G. D., "Models, Simulations, and Games--A Survey," Rand Report R-1060-ARPA/RC, The Rand Corporation, Santa Monica, CA, May 1972.
12. Greenberger, M. Crenson, M. A., and Crissey, B. L., Models in the Policy Process, Russell Sage Foundation, NY, 1976.
13. Gass, S. I., "Assessing Ways to Improve the Utility of Large-Scale Models," in [29].
14. Brewer, G., "What Ever Happened to Professionalism?" Interfaces, Vol. 8, No. 4, August 1978.
15. U. S. General Accounting Office, "Ways to Improve Management of Federally Funded Computerized Models," LCD-75-111, U. S. GAO, Washington, DC, 8/23/76.

16. Anon., "Activities of the Office of Energy Information and Analysis," Professional Audit Review Team, U. S. GAO, Washington, DC, 12/77.
17. Brewer, G., Politicians, Bureaucrats, and the Consultant--A Critique of Urban Problem Solving, Basic Books, NY, 1973.
18. Steele, J. L., The Use of Econometric Models by Federal Regulatory Agencies, Health-Lexington Books, NY, 1971.
19. Anon., Environmental Modeling and Decision Making, Report by the Holcomb Research Institute, 1976.
20. Pugh, R. E., Evaluation of Policy Simulation Models, Information Resources Press, Washington, DC, 1977.
21. Hogan, W. W., Sweeney, J. L., and Wagner, M. H., "Energy Policy Models in the National Energy Outlook," TIMS Studies in the Management Sciences, Elsevier North Holland, 1978.
22. Hogan, W. W., "Energy Modeling: Building Understanding for Better Use," Second Lawrence Symposium on Systems and Decision Sciences, 1978.
23. Hansman, J. A., "Project Independence Report: An Appraisal of U. S. Energy Needs Up to 1985," Bell Journal of Economics and Management, Vol. 6, 1975.
24. Isaacs, R., "On Applied Mathematics," Journal of Optimization Theory and Application, Vol. 27, No. 1, January 1979.
25. Anon., "Activities of the Energy Information Administration," Professional Audit Review Team, U. S. GAO, Washington, DC, 5/79.
26. Kuh, E., and Wood, D. O., "Independent Assessment of Energy Policy Models," Final Report, Research Project 1015-1, MIT Energy Laboratory, Cambridge, MA, 5/79.
27. DeMilo, R. A., Lipton, R. J., and Perlis, A. J., "Social Processes and Proofs of Theorems and Programs," Communication of the ACM, Vol. 22, No. 5, 5/79.
28. Gass, S. I., "Computer Model Documentation: A Review and an Approach," National Bureau of Standards, Special Publication 500-39, Washington, DC, 2/79.
29. Gass, S. I. (Editor), "Proceedings of the DOE/NBS Workshop on Validation and Assessment Issues of Energy Models," January 1979, National Bureau of Standards Special Publication, National Bureau of Standards, Washington, DC, (in press).
30. Greenberg, H., "The FEA Project Independence Experience," in [31].
31. Gass, S. I. (Editor), "Utility and Use of Large-Scale Mathematical Models," National Bureau of Standards, Special Publication 534, Washington, DC, 5/79.

32. Kresge, D., "An Approach to Independent Model Assessment," in [29].
33. Mulvey, J., "Strategies in Model Management," in [31].
34. Holloway, M., "The Texas National Energy Modeling Project: An Evaluation of EIA's Midrange Energy Forecasting System," in [29].
35. Gass, S. I., "Validation and Assessment of Energy Issues," paper presented at the NATO Research Institute, November 12-16, 1979, Brookhaven National Laboratory, Upton, NY.
36. Forrester, J. W., "Confidence in Models of Social Behavior--With Emphasis on System Dynamics Models," D-1967, System Dynamics Group, Sloan School of Management, MIT, Cambridge, MA, 1973.
37. Crissey, B., "Models in the Policy Process: A Framework," in [31].
38. Sargent, R. G., "Validation of Simulation Models," Working Paper #79-003, Department of Industrial Engineering and Operations Research, Syracuse University, Syracuse, NY, September 1979.
39. Mitchell, T. J. and Wilson, D. G., "Energy Model Validation: Initial Perceptions of the Process," ORNL/CSD-50, Oak Ridge National Laboratory, Oak Ridge, TN, November 1979.
40. Gass, S. I. et al, "Interim Report on Model Assessment Methodology: Documentation Assessment," NBSIR 80-1971, January 1980.
41. Greenberger, M., "Humanizing Policy Analysis: Confronting the Paradox in Energy Policy Modeling," paper to be presented at the Symposium on Model Validation and Assessment, May 1980.

APPENDIX A:

SURVEY RESULTS

As a follow-up to the Workshop on Model Confidence (October 4, 1979), we asked the participants to respond to a survey. The survey was intended to elicit opinions on the nature, importance, and feasibility of measuring model confidence. The results are presented in attachments I-1 and I-2. The same survey form was used in a similar request to the attendees at the NATO Brookhaven Energy Conference, November 10-14, 1979. The results are presented in attachments II-1 and II-2. The combined totals are given in attachments III-1 and III-2.

Conclusions reached from this type of ad hoc survey are usually difficult to justify in a statistical sense. Also, in rereading the seven statements, we perceive more ambiguity than we would have liked in a survey "instrument." However, it should be emphasized that the attendees of both the NBS Workshop and NATO Brookhaven Conference represent recognized expertise in energy modeling specifically, and in modeling in general. Their interpretation of the questions and their responses should be given much weight. Based on our conversations with the NATO attendees, it appears as if the European modeling community has not been concerned greatly with the concept of model confidence. Also, model validity is seen as a special concern of the modeler, but not of the decision maker. On the other hand, the Europeans indicate that their mode of operation tends to involve the decision maker much more than in the U. S., i.e., they claim that the decision maker is part of the modeling team. We have no other evidence that this is their standard practice in Europe. One can see from the surveys that there is a difference of opinion between the NBS (U. S.) and NATO (U. S. and European) groups. We give our interpretation of the responses by item. Some totals do not balance as a few respondents did not vote in all areas.

Respondents were asked to indicate a level of agreement with a set of statements on a discrete scale from -3 (strongly disagree) to +3 (strongly agree).

1. An operational definition and measures of model confidence can be developed that would be meaningful and of value to the model analyst (model developer).

	+3	+2	+1	0	-1	-2	-3
	-----			---	-----		
NBS		9		1		1	
NATO		<u>13</u>		<u>1</u>		<u>3</u>	
Combined		22		2		4	

It is clear that most respondents (who are modelers) believe that the concept of model confidence can be developed and prove to be of value to the modeling community.

2. An operational definition and measures of model confidence can be developed that would be meaningful and of value to the model user (policy maker).

	+3	+2	+1	0	-1	-2	-3
	-----			---	-----		
NBS		8		2		1	
NATO		<u>7</u>		<u>2</u>		<u>8</u>	
Combined		15		4		9	

In contrast to item 1, there is a fairly strong lack of consensus between the two respondent groups, with the NATO group being at most lukewarm about the prospects for meaningful measurement of confidence for decision makers. Note that in our report we stress that confidence is the decision maker's evaluation, not the modeler's.

3. A basic research problem in the development of an operational definition of model confidence is our being able to determine how the analyst's measures of confidence relates to the policy-maker's measures of confidence.

	+3	+2	+1	0	-1	-2	-3
	-----			---	-----		
NBS		7		1		3	
NATO		<u>6</u>		<u>6</u>		<u>5</u>	
Combined		13		7		8	

There does not appear to be much information in these scores. The intent of the item was to see if there was much difference in how the respondents viewed the two different concepts of confidence. Probably a poorly worded item.

4. The analyst and/or policy-maker measures for a specific model can be developed irrespective of competing models.

	+3	+2	+1	0	-1	-2	-3
	-----			---	-----		
NBS		6		2		2	
NATO		<u>5</u>		<u>1</u>		<u>7</u>	
Combined		11		3		9	

Based on comments, this item was the least understood. The question was intended to distinguish measures that could be applied independently to a single model from those which would be meaningful only in terms of comparisons, e.g., the former could possibly apply to a decision maker's mental model. Most agreement was in the categories of agree and mildly agree. The NATO group had a more or less balanced vote.

5. There is no value to DOE in furthering research on the topic of model confidence.

	+3	+2	+1	0	-1	-2	-3
	-----			---	-----		
NBS		0		0		11	
NATO		<u>1</u>		<u>2</u>		<u>14</u>	
Combined		1		2		25	

Because the item was worded "no" instead of "little" or "some," the respondents were more or less forced to disagree with it. But the overwhelming disagreement and the large number of strongly disagree votes indicates that research in the area is considered to be of value.

6. For most policy models, it is impossible to separate the model from the model analyst.

	+3	+2	+1	0	-1	-2	-3
	-----			---	-----		
NBS		7		0		4	
NATO		<u>12</u>		<u>1</u>		<u>4</u>	
Combined		19		1		8	

Most respondents agree with the sense of the item, with the NATO group coming out in stronger agreement. The results of this item should be interpreted along with those in item 7. It is not clear how the results of items 6 and 7 can be consistent unless "model analyst" in 6 was not equated to "original developers" in 7.

7. A DOE modeling goal should be to have all its models usable independent of the original developers.

	+3	+2	+1	0	-1	-2	-3
	-----			---	-----		
NBS		11		0		0	
NATO		10		1		5	
Combined		<u>21</u>		<u>1</u>		<u>5</u>	

There was some question on this item due to the use of "all" instead of "some" or "most" or other qualifiers. The NBS group, that included DOE personnel and modelers and consultants for DOE, was in strong agreement (11 to 0). The NATO group, that also included some DOE and DOE consultant groups, was 2 to 1 in agreement. This item should be interpreted along with that of item 6. We can conclude that the respondents feel that the milieu of a model must include the modeler, but the user (here DOE) should attempt to separate its models from the original developers; independence does not rule out the active keeping of in-house or consultant analysts for the models.

APPENDIX B:

RESPONDENTS FROM NATIONAL BUREAU OF STANDARDS WORKSHOP



UNITED STATES DEPARTMENT OF COMMERCE
National Bureau of Standards
Washington, D.C. 20234

-27-

October 23, 1979

MEMORANDUM FOR Participants of the Model Confidence Workshop

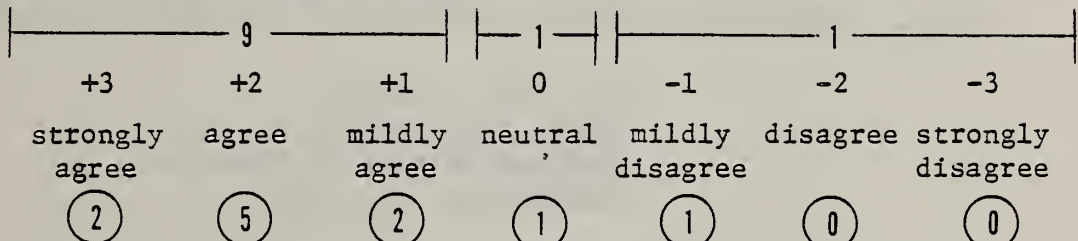
From: Saul I. Gass
Operations Research Division

Subject: Model Confidence Survey

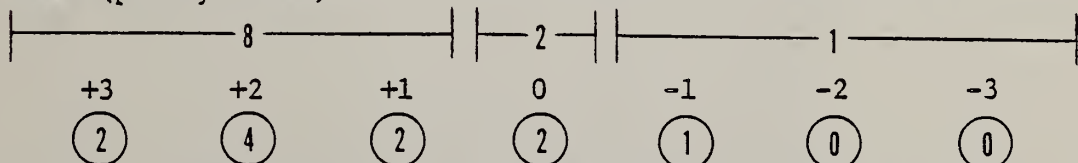
The following opinion survey is designed to obtain your views of significant issues relative to model confidence. Your completing and returning it within ten days would be appreciated. I will forward a summary of the results to each of you.

Please indicate your sense of agreement or disagreement by circling the appropriate number.

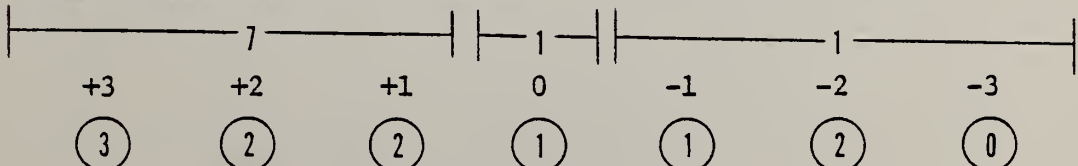
1. An operational definition and measures of model confidence can be developed that would be meaningful and of value to the model analyst (model developer).



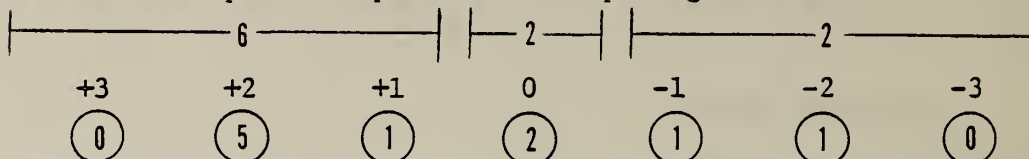
2. An operational definition and measures of model confidence can be developed that would be meaningful and of value to the model user (policy maker).



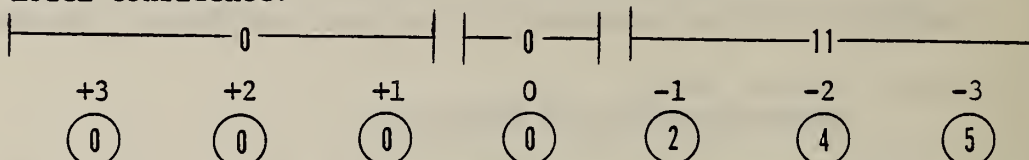
3. A basic research problem in the development of an operational definition of model confidence is our being able to determine how the analyst's measures of confidence relates to the policy-maker's measures of confidence.



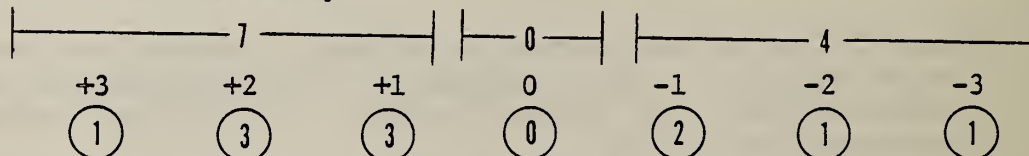
4. The analyst and/or policy-maker measures for a specific model can be developed irrespective of competing models.



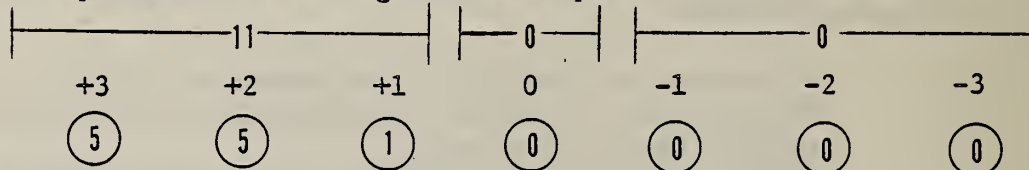
5. There is no value to DOE in furthering research on the topic of model confidence.



6. For most policy models, it is impossible to separate the model from the model analyst.



7. A DOE modeling goal should be to have all its models usable independent of the original developers.



Please return to:

Dr. Saul I. Gass
A428, Building 101
National Bureau of Standards
Washington, DC 20234

(signature)

APPENDIX C:

RESPONDENTS FROM THE NATO BROOKHAVEN ENERGY CONFERENCE



-29-

October 23, 1979

MEMORANDUM FOR Participants of the Model Confidence Workshop

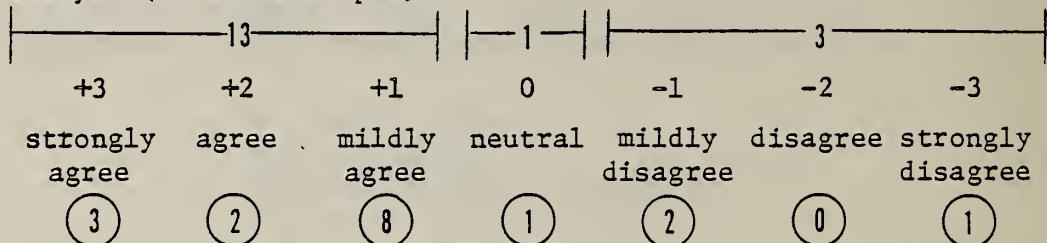
From: Saul I. Gass
Operations Research Division

Subject: Model Confidence Survey

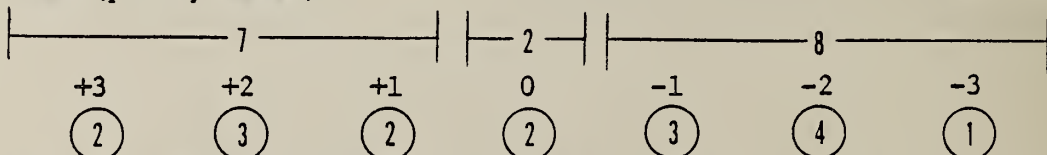
The following opinion survey is designed to obtain your views of significant issues relative to model confidence. Your completing and returning it within ten days would be appreciated. I will forward a summary of the results to each of you.

Please indicate your sense of agreement or disagreement by circling the appropriate number.

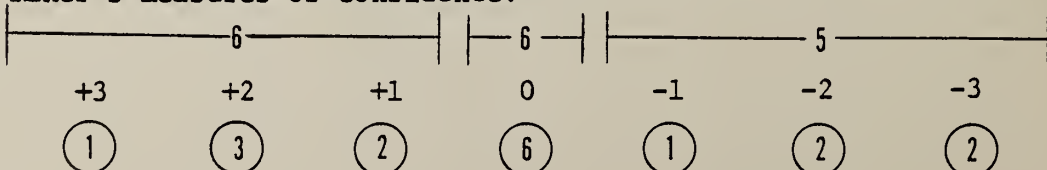
1. An operational definition and measures of model confidence can be developed that would be meaningful and of value to the model analyst (model developer).



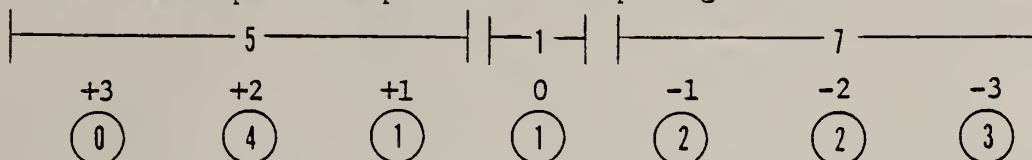
2. An operational definition and measures of model confidence can be developed that would be meaningful and of value to the model user (policy maker).



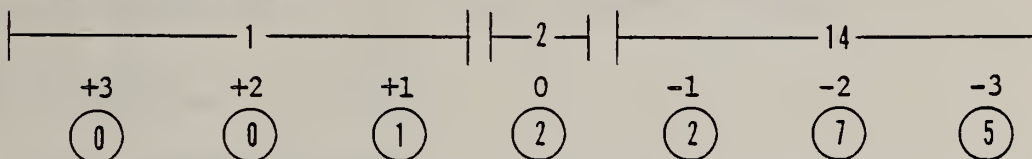
3. A basic research problem in the development of an operational definition of model confidence is our being able to determine how the analyst's measures of confidence relates to the policy-maker's measures of confidence.



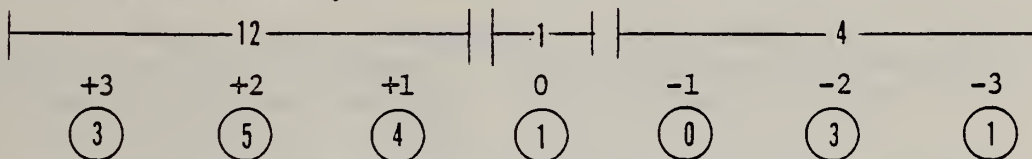
4. The analyst and/or policy-maker measures for a specific model can be developed irrespective of competing models.



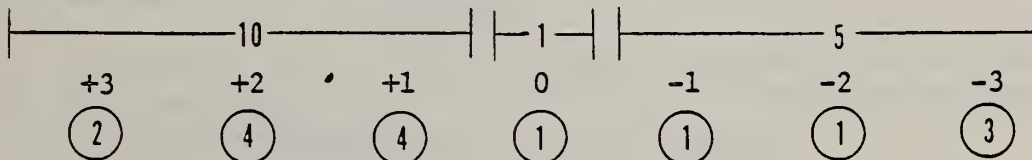
5. There is no value to DOE in furthering research on the topic of model confidence.



6. For most policy models, it is impossible to separate the model from the model analyst.



7. A DOE modeling goal should be to have all its models usable independent of the original developers.



Please return to:

Dr. Saul I. Gass
A428, Building 101
National Bureau of Standards
Washington, DC 20234

(signature)

APPENDIX D:

COMBINED NBS/BROOKHAVEN RESPONDENT TOTALS



UNITED STATES DEPARTMENT OF COMMERCE
National Bureau of Standards
Washington, D.C. 20234

-31-

October 23, 1979

MEMORANDUM FOR Participants of the Model Confidence Workshop

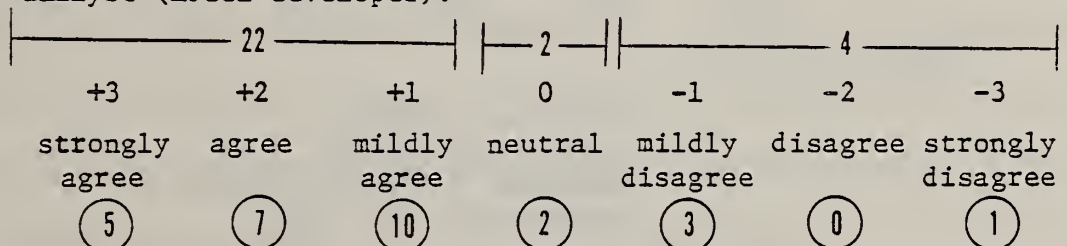
From: Saul I. Gass
Operations Research Division

Subject: Model Confidence Survey

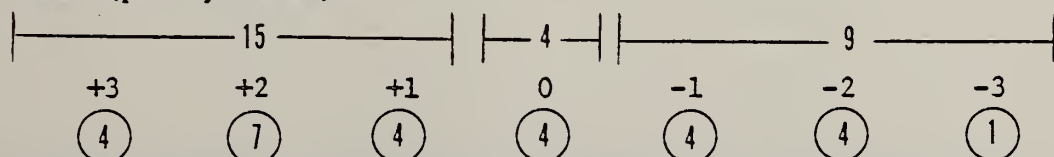
The following opinion survey is designed to obtain your views of significant issues relative to model confidence. Your completing and returning it within ten days would be appreciated. I will forward a summary of the results to each of you.

Please indicate your sense of agreement or disagreement by circling the appropriate number.

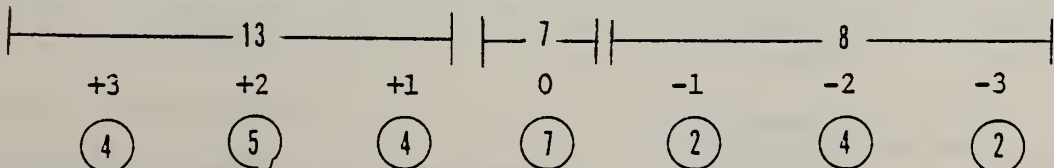
1. An operational definition and measures of model confidence can be developed that would be meaningful and of value to the model analyst (model developer).



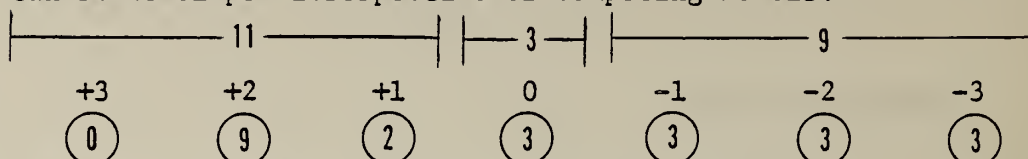
2. An operational definition and measures of model confidence can be developed that would be meaningful and of value to the model user (policy maker).



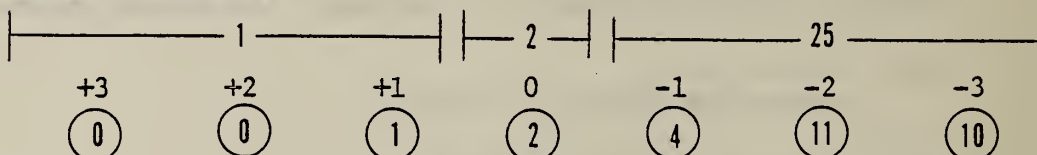
3. A basic research problem in the development of an operational definition of model confidence is our being able to determine how the analyst's measures of confidence relates to the policy-maker's measures of confidence.



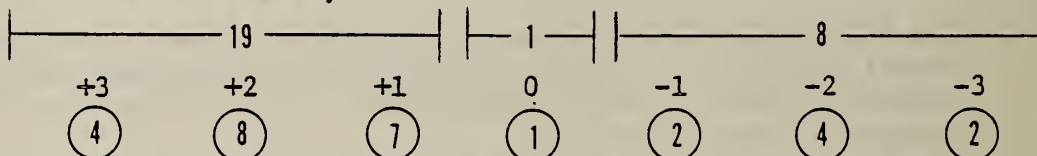
4. The analyst and/or policy-maker measures for a specific model can be developed irrespective of competing models.



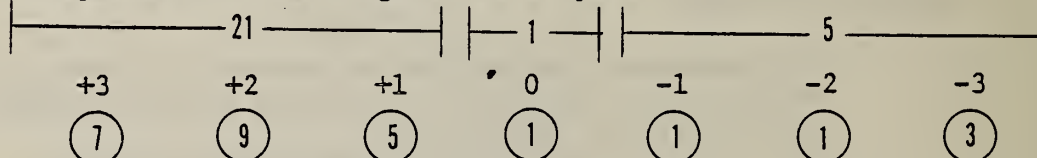
5. There is no value to DOE in furthering research on the topic of model confidence.



6. For most policy models, it is impossible to separate the model from the model analyst.



7. A DOE modeling goal should be to have all its models usable independent of the original developers.

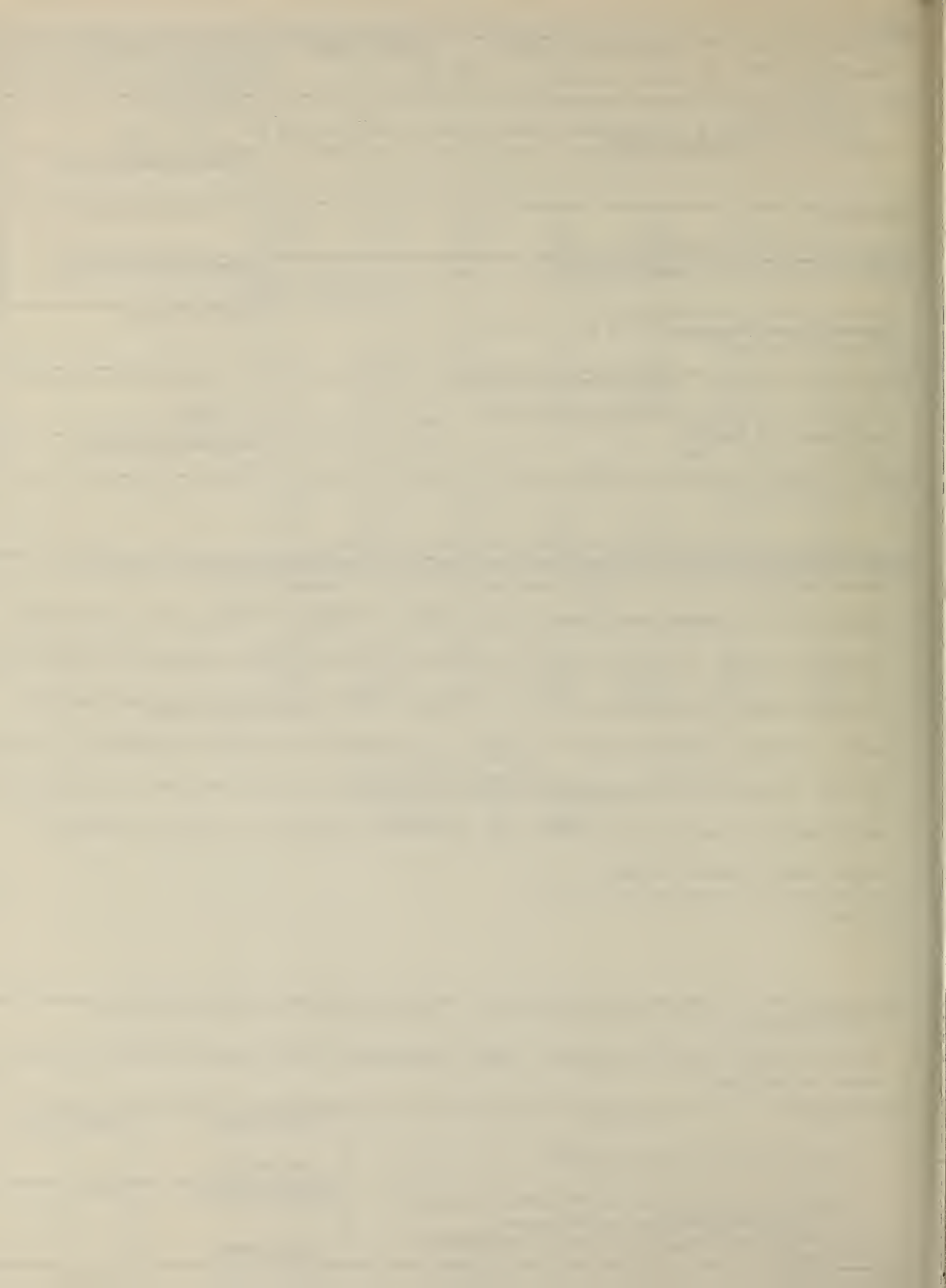


Please return to:

Dr. Saul I. Gass
A428, Building 101
National Bureau of Standards
Washington, DC 20234

(signature)

U.S. DEPT. OF COMM. BIBLIOGRAPHIC DATA SHEET		1. PUBLICATION OR REPORT NO. NBSIR 80-2053	2. Gov't. Accession No.	3. Recipient's Accession No.
4. TITLE AND SUBTITLE Concepts of Model Confidence			5. Publication Date June 1980	
			6. Performing Organization Code	
7. AUTHOR(S) Saul I. Gass and Lambert S. Joel			8. Performing Organ. Report No.	
9. PERFORMING ORGANIZATION NAME AND ADDRESS NATIONAL BUREAU OF STANDARDS DEPARTMENT OF COMMERCE WASHINGTON, DC 20234			10. Project/Task/Work Unit No.	
			11. Contract/Grant No.	
12. SPONSORING ORGANIZATION NAME AND COMPLETE ADDRESS (Street, City, State, ZIP) Office of Oversight Analysis and Access Department of Energy Washington, DC 20461			13. Type of Report & Period Covered NBSIR	
			14. Sponsoring Agency Code	
15. SUPPLEMENTARY NOTES <input type="checkbox"/> Document describes a computer program; SF-185, FIPS Software Summary, is attached.				
16. ABSTRACT (A 200-word or less factual summary of most significant information. If document includes a significant bibliography or literature survey, mention it here.) This report discusses the concept of confidence in results obtained from large-scale modeling systems. It is written in satisfaction of the "model confidence" tasks of a National Bureau of Standards project on "Energy Model Validation Procedure Development," funded by the Department of Energy. This report includes discussions of: our efforts to define model confidence; the workshop held for this purpose; a preliminary methodology to measure confidence; and, the survey conducted to obtain opinions on significant related issues.				
17. KEY WORDS (six to twelve entries; alphabetical order; capitalize only the first letter of the first key word unless a proper name; separated by semicolons) Decision making; model assessment; model confidence; model evaluation; model utility; model validation.				
18. AVAILABILITY <input checked="" type="checkbox"/> Unlimited <input type="checkbox"/> For Official Distribution. Do Not Release to NTIS <input type="checkbox"/> Order From Sup. of Doc., U.S. Government Printing Office, Washington, DC 20402, SD Stock No. SN003-003- <input checked="" type="checkbox"/> Order From National Technical Information Service (NTIS), Springfield, VA, 22161		19. SECURITY CLASS (THIS REPORT) UNCLASSIFIED		21. NO. OF PRINTED PAGES 40
		20. SECURITY CLASS (THIS PAGE) UNCLASSIFIED		22. Price \$6.00





NBSIR 80-2053

- -

- -